TECHNICAL STANDARD OPERATING PROCEDURE

Date: May 16, 2001 SOP No. WGI-VBI70-01

Title: **Electronic Data Upload**

APPROVALS:

Washington Group International, Inc.

Author: Don Bott, Marta Green Date: May 15, 2001

SYNOPSIS: This Standard Operating Procedure (SOP) describes the processes used to incorporate information received in digital format from the participating laboratories into the project database.

Received by QA Unit

REVIEWS:

TEAM MEMBER	SIGNATURE/TITLE	DATE
EPA Region 8		
Washington Group		5/16/01

REV.	DATE	REVISION DESCRIPTION

TECHNICAL STANDARD OPERATING PROCEDURE
ELECTRONIC DATA UPLOAD

1.0 PURPOSE

The purpose of this SOP is to outline the process by which electronic analytical data are uploaded and reviewed consistent with the following objectives:

- Effect the transfer of laboratory analytical results to the project database with full admissibility checking and minimal user effort;
- Prevent laboratory source data files from being processed more than once;
- Identify inadmissible sample records, and suggest / allow corrective action on these sample records without changing the original data files;
- Prevent duplicate sample information from being uploaded into the project database;
- Provide audit records of the results.

2.0 SCOPE

This Standard Operating Procedure (SOP) describes the processes used to incorporate information received in digital format from the participating laboratories into the project database.

3.0 RESPONSIBILITIES

Individual and organizational responsibilities for data management personnel are described in the Data Management Plan.

The Field Activities Database Manager is responsible for overseeing the accurate and complete population and maintenance of the computerized database used to electronically store and process data obtained during field collection activities. The Field Activities Database Manager is responsible for electronic data uploading, including posting of data to staging files, reviewing data integrity, modifying electronic data as appropriate, and loading the data to the project database. The Field Activities Database Manager is also responsible for electronic database and document security.

TECHNICAL STANDARD OPERATING PROCEDURE
ELECTRONIC DATA UPLOAD

The Project Chemist responsible for verification of electronic deliverables may assist in the data review, identification of data discrepancies, and data modification or qualification.

4.0 PROCESS OVERVIEW

The process steps include:

- transfer of information from the data files received from the laboratories into staging / review tables in the database;
- facilities for review of a read-only report of the laboratory information;
- corrections or amendments of data in accord with Section 5.8.3;
- transfer of validated laboratory information from the staging / review tables to the permanent storage tables in the database.

The upload process is implemented and controlled by the upload software application specifically designed for this purpose. This application is designed so that the user need not perform any sophisticated file editing or database operations.

5.0 TRANSFER OF DATA FILE CONTENTS TO STAGING TABLES

Upon receipt of one or more data files to be transferred to the project database, a new folder will be created under the main project folder. The folder name indicates that the data files are to be uploaded, in **LOAD**YYYYMMDD format. For example, if the upload date is June 15, 2001, the folder name would be LOAD20010615.

The incoming files are transferred to this folder. To avoid inadvertent alteration of the files, Windows Explorer should be used to change the Properties of each file so the "Read-only" attribute is set (box is checked in the Properties tab). A randomly selected set of the files will be reviewed to ensure that the file format is consistent with the lab's prescribed format. No changes or modifications may be made to the input files at any time.

The first phase of the upload application software displays the available folders to the user, who selects the folder to be processed. Each file in the folder will be read and transferred as received to a staging table in the database. If there are problems transferring a file's contents, the file is identified to the user and processing continues with the next file. The user is responsible for follow-up action for any file that cannot be processed through this step. At the end of this phase the list of files and upload status is displayed on a report for review / printing.

The second phase of the upload application software adds identifying information including the

file name, date created, and date received, and then performs admissibility checks on the laboratory information. Processing at this step includes but is not limited to:

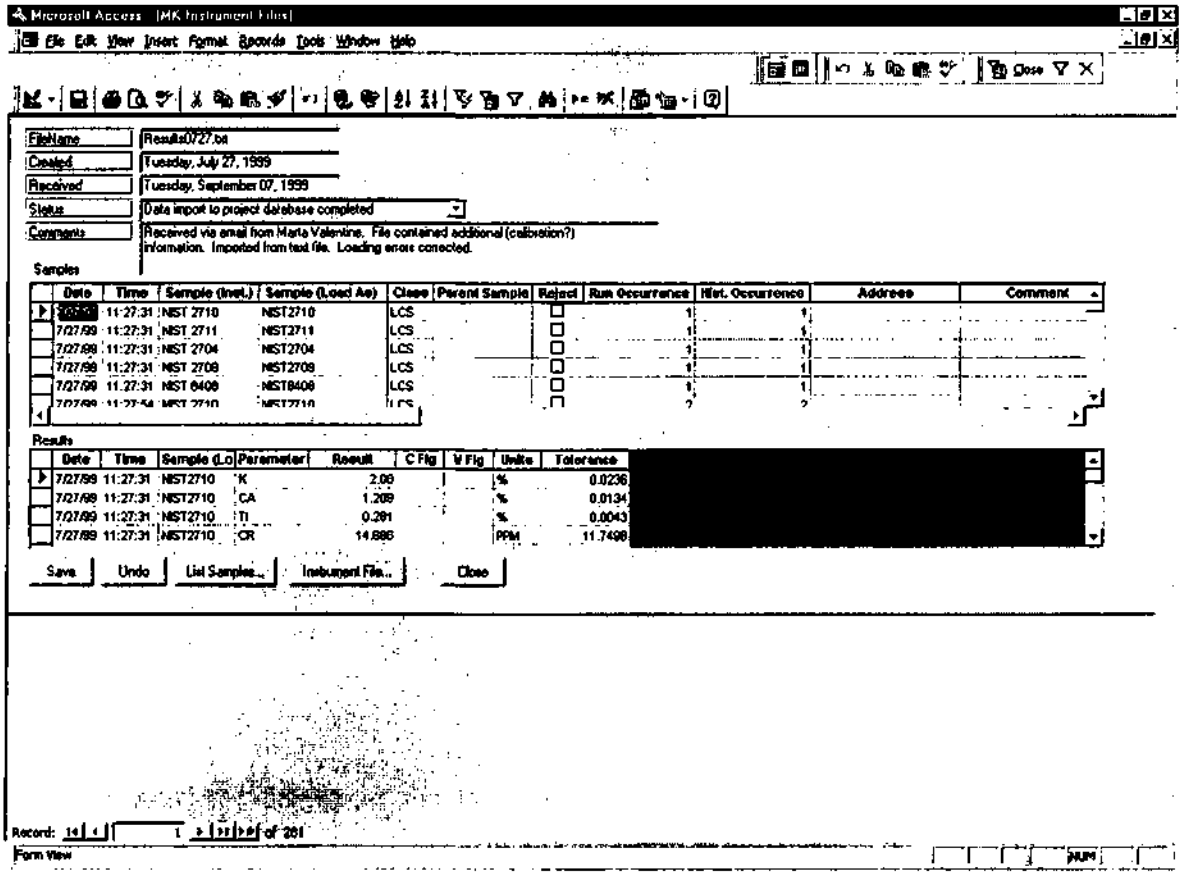
- checking each sample ID against the Field Sampling information table to identify the field sample for lab samples;
- checking the analyte against the analyte list for the project;
- identifying the sample "class" (e.g., Laboratory Control Sample, Field Sample);
- checking for historical occurrences of the same sample ID;
- identifying the property where the field sample is taken, if available, and flagging those samples without a matching property.

6.0 REVIEW OF THE STAGED DATA FILES AND CORRECTIVE ACTION AS REQUIRED

At the end of the second phase, the user will review and print reports of the sample and associated analytical information. The data review is achieved by selecting Data Maintenance from the Main Menu, then Laboratory Data Entry and Review Files for Upload. The file(s) of interest are selected by filtering for date, file name, or file status in the data screen header. The data displayed for review are:

- Date - of analysis
- Time - of analysis
- Sample (Instr.) – Sample ID as reported by the instrument
- Sample (Load as) – Identical to Sample (Instr.) unless modified
- Class – sample class
- Parent Sample – Parent Sample ID, blank unless modified
- Reject – toggle to reject all associated analytical results for that Sample ID
- Run Occurrence – number of occurrences of the Sample ID in that data file
- Hist. Occurrence – number of occurrences of the Sample ID total in the file and previously loaded into the database
- Address – property address if available in the database from field sampling data entry
- Comment – to document reason for data modification or qualification

An example screen is shown below:



The failure modes for samples failing the admissibility criteria will be displayed. The admissibility criteria can be found in Attachment 1. Additional discrepancies that requires data modification or qualification are indicated by the following conditions:

- Historical occurrence (equal to two or greater occurrences) of the Sample ID for a Field Sample
- No Property Address for a Field Sample

Historical occurrence should not exist for a field sample (Class=FS) unless the sample was re-analyzed. Only one result is to be reported for each Sample ID, and additional (previously or currently under review) results must be qualified as Rejected. A determination of the final data versus that which is to be rejected is made by the Project Chemist based on laboratory documentation. Prior to uploading, the Reject toggle on the review screen may be used to qualify all results for that sample analysis.

If a Property Address is not located for a field sample (Class=FS), then generally the Sample ID

is not actually a field sample but rather a blind QC sample. Alternatively, a typographical error may exist either in the field sample ID already in the database from the sampling data entry, or in the laboratory sample ID entered in the laboratory. The Project Chemist resolves this discrepancy by reviewing QC sample logs (blind split samples, blind standards). QC samples are identified in the staging table by changing the Class code (i.e., BS for Blind Split) and entering the parent sample ID listed in the QC sample log.

If a field sample is not matched to a property address and is determined not to be a QC sample, the Project Chemist reviews the Sample Preparation Log and other field and laboratory documents as needed to resolve the discrepancy. If a typographical error is identified, the Sample ID is changed in the Sample (Load as) field.

Any modification of a Sample ID or rejection of data requires documentation under the Comments field, including the reason for the change. The failure mode for each sample in the staging table is checked when the user has saved edits / changes; if the admissibility criteria are met, then the sample is marked as admissible. Samples that do not meet admissibility criteria will remain in the staging tables until corrected. When the review is complete, the user changes the file Status to Review Complete, Ready for Upload.

7.0 TRANSFER OF SAMPLE INFORMATION FROM THE STAGING TABLES TO THE PROJECT DATABASE

The final phase in the process is the transfer of information from the staging tables to the project database. As noted above, only admissible samples are eligible for transfer.

In addition, each candidate laboratory sample number is checked against the existing laboratory sample number in the project database. If the sample is already in the project database, then the candidate sample is flagged as a duplicate sample and is not transferred from the staging table.

If the sample passes this check, the contents of the sample record are added to the corresponding table in the project database. The status flag for the staging table record is set to show if the transfer is successful or unsuccessful. A typical reason for failure in the transfer is loss of connectivity to the database server.

At the end of this phase, the user will review the transfer report showing

- the lab file folder and file name;
- the date and time processed;
- the number of samples transferred;
- the number of samples eligible but not transferred due to communications failures, etc.;
- the number of duplicate samples;
- the number of inadmissible samples remaining in the staging table.

8.0 AUDIT RECORD OF UPLOAD ACTIVITY

The transfer activity report will be printed and maintained in a project folder for review as required. In addition, the upload software application can produce reports showing

- the files and associated processing dates and times;
- the inadmissible samples remaining to be corrected and uploaded;
- the eligible samples awaiting transfer to the project database;
- the samples whose ID numbers are duplicates of those already in the database.

9.0 MAINTENANCE OF THE STAGING TABLES

The staging tables will be maintained to provide a permanent record of the review process and data modifications and qualifications made through implementation of this procedure.

ATTACHMENT 1

ADMISSIBILITY CRITERIA

The following addresses the admissibility criteria for data received in Paragon Laboratory format to be uploaded into the Vasquez Boulevard - I-70 (VBI70) project database. Except where noted, these admissibility criteria will be enforced by the upload software application.

The methodology employs "staging tables" for the input file names and the information contained in these files. These staging tables are for the use of the upload application only. No laboratory data enters the project database without passing through the upload application.

Admissibility Criterion	Method of Enforcement
How do we ensure that the same file is not processed twice?	1. The uniqueness of the file names in a folder is enforced by the Windows file system. 2. The folder / file names are checked against the history of previously processed file names by the upload application. If the same folder / file name is submitted to the upload application more than once, the subsequent occurrences are ignored.
What kinds of "lab files" can be processed?	1. The file name must end in ".TXT" 2. The file data must conform to the Paragon Laboratory format, with (a) the first (header) row consisting of the field names, and (b) subsequent records containing the corresponding data items in comma-delimited format. If either of these checks fail, the status flag for corresponding staging table record for the file is set accordingly.
Once the lab data has been loaded into the staging table, what kinds of checks are performed?	1. The analyte is checked against the reference list for analytes. 2. The "QC Type" field is checked against

	the reference list. Typical entries are "field sample", "laboratory duplicate", and "lab matrix spike duplicate". 3. If the "Field ID" is present, it is checked against the current list of field sample IDs in the project database. 4. If the "Field ID" is valid, the associated property identification in the "Site Visits" project database table is checked against the property list. These checks are performed in the order shown. If any of the checks fails, the Sample Status flag is set to identify the failure mode.
How do we ensure that inadmissible data are not moved from the staging tables to the project database?	No staging table records can be moved unless the status code indicates all problems have been corrected.
How will the project team review / correct information?	1. The update application can produce complete reports on the data entered into the staging tables at any point. 2. The project team can review and edit the staging table data using the upload application.
How can we ensure that the editing process does not introduce new problems?	The project team cannot change the status flag for any samples. The upload application has a "Revalidate" procedure to check each sample for admissibility criteria after any edits are made.
How long will inadmissible data remain in the staging tables?	The records remain until 1. all admissibility criteria are met; or 2. the project team executes an upload application process that copies the

	appropriate records into an archive table and then removes them from the staging tables..
Can duplicate records be moved from the staging tables to the project database?	The sample ID in the staging table record is checked against the sample IDs in the project database. A staging table record with a duplicate ID is flagged and is not loaded into the project database. A special procedure allows the bulk removal of these duplicate sample IDs from the staging table, since their only inadmissibility fault is that they were processed more than once.
How can we identify the staging table records that have been successfully transferred?	Each sample ID in the staging table that should have been transferred is validated against the project database. This guards against data transfer loss due to communications failures, etc.
How do we keep the staging tables from growing without limit?	Through the upload application, the project team can remove admissible and transferred samples from the staging tables at any time.